Finding Real-life Doppelgangers on Campus with MTCNN and CNN-based Face Recognition

Jingjing Ye
ShanghaiTech University
393 Huaxia Middle Rd, Pudong Xinqu
Shanghai, China 201210
yejj@shanghaitech.edu.cn

Yilu Zhou
Fordham University
140 W 62nd St
New York, NY 10023, U.S.A.
yzhou62@fordham.edu

Abstract

Face recognition has many applications such as national security, finding missing people, forensic investigation, and face payment. Inspired by Francois Brunelle's doppelganger project, where he spent 12 years tracking people who are completely strangers but lookalikes, this study aims to use face recognition techniques to mine doppelgangers on campus. Face Detection and Face Recognition has been widely used in public security and the techniques are fast evolving. There is a great potential to combine technology and art activities, in this case mimicking a photographers' human eyes. We develop a face processing system which includes four steps, face detection, image processing (alignment, cropping), feature extraction and classification. The Multi-task Cascaded Convolutional Networks (MTCNN) is used to detect faces and traditional CNNs with Softmax Loss and Center Loss joint is used to train on a Caffe framework. Finally, cosine similarity is used to calculate the eigenvector of two faces and derive their face similarity. The outcome of this study contributes to the application of face recognition with real-life data and provides possible collaboration channels between technology and art.

Keywords: doppelganger, face recognition, face detection, MTCNN, CNNs, softmax, center loss, cosine similarity

1 Introduction

In the past, technology and art are not often associated together. However, with the rapid development in computer vision, there are more and more interest to combine technology with art, especially in the area of image generation. For example, in 2016 a team of technologists produced a 3D-printed painting in the style of Dutch master Rembrandt (https://www.nextrembrandt.com/). Previous paintings of Rembrandt were used to calculate distances between facial features. In 2018, an art painting generated using an algorithm and a data set of 15,000 portraits painted between the 14th and 20th centuries was sold for $432,500 (Time, 2018). Such activities have inspired further interest among computer scientists to look for new applications of computer algorithms in art world. It also encourages artists to seek new technology as an expression of art.

We are inspired by photographer François Brunelle's work on "Me, Myself and I" where he spent 12 years running around many countries and cities to find couples who are complete strangers but look like each other, or doppelgangers in real life (Chase Jarvis Photography, 2018). His art exhibition attracted many people. Figure 1 shows some couples he photographed for the exhibition. Following the work of François Brunelle, this study aims to mine doppelgangers on a university campus using face detection and face recognition algorithms.

There are several potential contributions of this study. Firstly, the study will demonstrate the potential of using technology in creative art. We show that technology can be more accurate in capturing similar faces and eliminate human visual bias affected by skin colors and gender. Secondly, the result of this study will be displayed as an exhibition on campus following Brunel's exhibition idea. Furthermore, there will be interactive activities that allow visitors find their similar face on campus and test their face similarity with friends. This exhibition will include introduction of recent development in Deep Learning and Image Processing technology. It will inspire students in non-computer fields to explore the possibility of cross-disciplinary research. This will especially motivate students in art-related major to explore new ways of art expression. It will contribute to the area of face recognition by using a real-life dataset of over 3,000 portraits. And lastly, we show the algorithm's potential contribution to e-business in finding illegal use of celebrities' portraits by altering the threshold. Specifically, we show the importance of face alignment in CNN-based face recognition. Technologies developed in this study can be applied in a variety of applications.

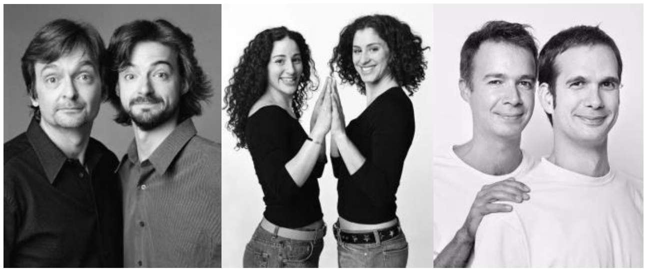


Figure 1. Photos from Brunel Photography's Exhibition (Chase Jarvis Photography, 2018)

2 Literature review

Face recognition has experienced rapid development in recent years with deep learning infrastructure. With proper training, computer algorithms may perform as well as or even better than human experts in identifying similar faces. These face recognition algorithms provide foundations for our study.

Many research papers use common dataset to test their algorithms. For example, the CMU Multi-PIE Face Database contains more than 750,000 images of 337 people (http://www.multipie.org/). The CAS-PEAL face database contains 99,594 images of 1040 Chinese individuals (595 males and 445 females) (http://www.jdl.ac.cn/peal/index.html). The 10k US Adult Faces Database is a collection of 10,168 natural face photographs for 2,222 of the faces (http://www.wilmabainbridge.com/facememorability2.html). While these datasets provide great resources for face recognition studies, algorithms are often trained for a particular dataset and the generalizability is unknown. Thus, a real-life face recognition dataset will contribute to the validity of current algorithms.

Face detection is the first important step of face recognition. A classic and simple method of face detection is Haar cascaded and AdaBoost algorithms. It was first proposed by Viola and Jones for rapid object detection (2001). It allows rapid object detection using a boosted cascaded of simple features with an accuracy of above 80% (Viola and Jones, 2004). CNN-based face detection has been a mainstream approach for object detection and recognition in recent years. AlexNet was a pioneer in applying CNN-based object detection and won the ImageNet Large Scale Visual

Recognition Challenge (Krizhevsky et al., 2012). Zhang et al. (2014) invented a Tasks-Constrained Deep Convolutional Network(TCDCN) to detect bounding boxes of human faces and face landmarks. TCDCN achieved higher detection accuracy than Cascaded CNN and the approach also yielded a significantly lower computational cost. MTCNN is a more recent CNN-based model that combines deep multi-task learning and cascaded networks. Its performance is better than TCDCN and Cascaded CNN alone. Besides MTCNN, there are some special face detection algorithms such as Gabor filter (Suri et al. 2011) which performs well under different light effect. But for a generic scenario, CNN-based detections are still the most popular.

Face recognition phase aims to extract important features on faces. This is done by performing supervised learning in CNN with an appropriate loss function. The identification of an ideal loss function is critical to the learning performance. For example, Softmax loss function can generate rich face features for recognition. Researchers often combine multiple loss functions. For example, Smirnov et al. (2017) combined Softmax loss with Margin-based loss. The final loss function is L2S + λMB. (λ is the ratio). By tuning the ratio, one can achieve a higher performance by joint supervision. There are several challenges in face recognition. Different poses (Logie et al. 1987), lightings (Georghiades et al. 2001), expressions (Sim et al. 2002) and occlusions (such as glasses) will affect face recognition seriously. Thus, standardization of images is a necessary step. This step is called alignment (Taigman et al. 2014).

Finally, there are few studies look into doppelganger problem. Smirnov et al. (2017) proposed doppelganger mining based on CMU's Multi-PIE database of 337 people faces, a much smaller dataset compare to ours. We are interested in applying face detection and face recognition algorithms in finding real-life doppelgangers on a university campus in China. The technical focus of the study is the integration of available techniques and the specific challenges dealing with this new dataset.

3 Proposed Methodology

In order to find lookalike faces, or doppelgangers, we proposed a four-phase framework that consists of: (1) face detection, (2) face alignment, (3) face recognition and (4) similarity computation as shown in Figure 2. Face detection component identifies the position of the face. Face alignment component rotates the face to allow cross-image comparison. Face recognition trains a feature extraction model and finally similarity computation is performed. We explain each component in this section.

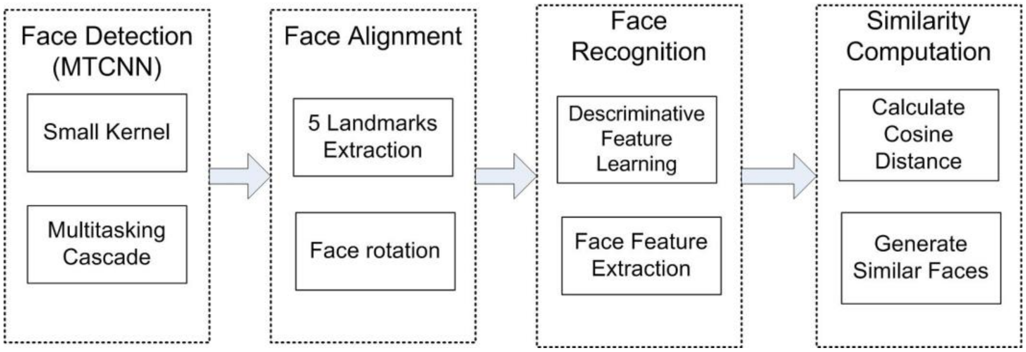


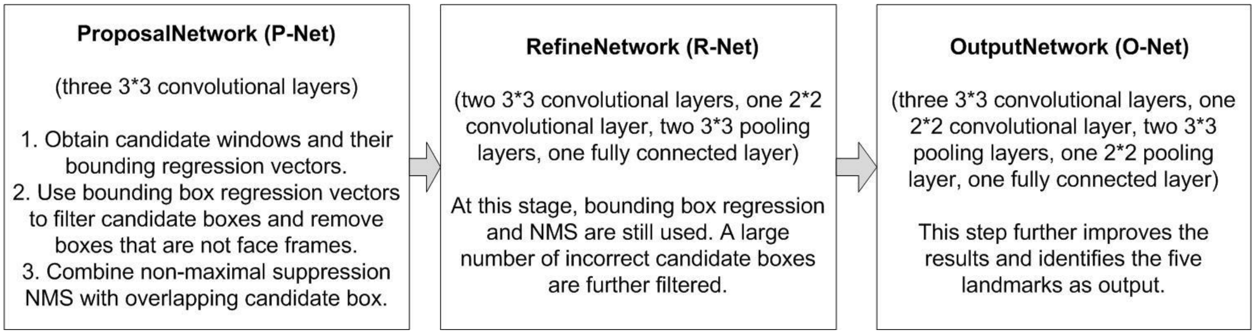
Figure 2. Framework of Doppelganger Mining

3.1 Face Detection (MTCNN) and Alignment

The purpose of face detection is to extract face features, or landmarks from the image of face. Based on our review, we chose to use a Multitasking Convolutional Neural Network (MTCNN) (Zhang et. al., 2016). The algorithm is effective and combines the advantage of deep multi-task learning and cascaded network. MTCNN marks five landmarks: two eyes, tip of the nose, and two corners of mouth.

We used a small convolution kernel (size of 3*3 and 2*2) in our neural network which has several advantages in face detection. It enables more hidden layers and more nonlinear functions, improves the discretion of decision function and reduces the number of parameters to be trained (Simonyan et al., 2014). Because MTCNN contains multiple cascaded networks, number of parameters will greatly slow down training performance.

MTCNN is a sequence of network where the output from the previous network is passed on to the input of the next network. Similar to AdaBoost algorithm, MTCNN iterates through a sequence of weak classifier to generate a strong classifier to detect five accurate landmarks. Following Zhang's work (2016), we use three networks to generate accurate landmarks: Proposal Network (P-Net), Refine Network (R-Net) and Output Network (O-Net) and the process is illustrated in Figure 3.



ProposalNetwork (P-Net)

(three 3*3 convolutional layers)

1. Obtain candidate windows and their bounding regression vectors.
2. Use bounding box regression vectors to filter candidate boxes and remove boxes that are not face frames.
3. Combine non-maximal suppression NMS with overlapping candidate box.

RefineNetwork (R-Net)

(two 3*3 convolutional layers, one 2*2 convolutional layer, two 3*3 pooling layers, one fully connected layer)

At this stage, bounding box regression and NMS are still used. A large number of incorrect candidate boxes are further filtered.

OutputNetwork (O-Net)

(three 3*3 convolutional layers, one 2*2 convolutional layer, two 3*3 pooling layers, one 2*2 pooling layer, one fully connected layer)

This step further improves the results and identifies the five landmarks as output.

Figure 3. MTCNN algorithm with three Networks

MTCNN identifies three important descriptors for training: face classifier, face boundary box, and five landmarks. Face classifier identifies whether the object is a face or not. It can be achieved by a cross entropy loss function shown in formula (1). For each picture *i*, p_i is the probability of a human face is determined, where y_i^{det} is a background label that is marked in advance in a standard training set.

$$L_i^{det} = -(y_i^{det} log(p_i) + \left(1 - y_i^{det}\right)(1 \; - log(p_i)))$$
$$y_i^{det} \epsilon \{0,1\} \tag{1}$$

Once an object is classified as a face, face boundary box will be calculated based on Euclidean distance to identify Boundary box regression as shown in formula (2), where $\hat{y}$ is the results predicted by the network. y is the correct coordinates labeled previously, which is a quaternary:

[$x_{upper\ left\ corner\ of\ bounding\ box}$, $y_{upper\ left\ corner\ of\ bounding\ box}$,
long of the bounding box , wide of bounding box]

$$L_i^{box} = \left|\hat{y}_i^{box} - y_i^{box}\right|_2^2$$

$$y_i^{box} \epsilon R^4$$

(2)

Finally, five key points of the face, or landmarks are positioned as shown in formula (3). Here, $y_i^{landmark}$represents five landmark points, using a pair of values (x,y), where $\hat{y}$ is the predicted face landmarks, and *y* is the actual marked landmark.

$$L_i^{landmark} = \left|\hat{y}_i^{landmark} - y_i^{landmark}\right|_2^2$$
$$y_i^{landmark} \epsilon R^{10}$$

(3)

With above three descriptors, the supervised function of final training network is illustrated in formula (4). The goal is to minimize the Euclidean distance between the actual marks and the predicted marks. The result of the three networks is five predicted landmarks.

$$min \sum_{i=1}^{N} \sum_{j\in\{det,box,landmark\}} \alpha_j \beta_i^j l_i^j$$
$$\beta_i^j \epsilon \{0,1\}$$
$$\text{P-Net R-Net } (\alpha_{det} = 1 , \alpha_{box} = 0.5, \alpha_{landmark} = 0.5)$$
$$\text{O-Net } (\alpha_{det} = 1, \alpha_{box} = 0.5, \alpha_{landmark} = 1)$$

(4)

N: number of training samples
α_j : the importance of this loss function
β_i : sample label
Li^j: one of three loss functions mentioned before

3.2 Face Alignment

After the detection of 5 predicted landmark points, the pictures need to be rotated and resized for similarity comparison in the final step. We first aggregate all pictures and generate a group of mean landmarks. This is used as the standard to align all pictures. The five landmarks are then positioned to the closed position to the mean standard. Figure 4 shows the process of (1) detecting 5 landmarks and (2) face alignment and cropping. The tilted face in the original picture on the left is cropped and becomes correctitude.

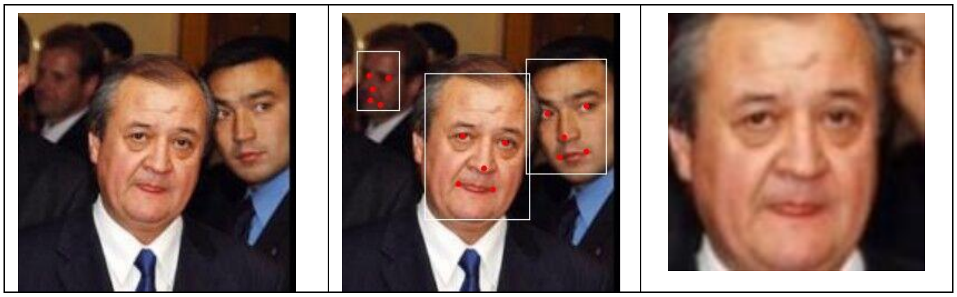


Figure 4. (from left to right) Original picture, Face Detection, Face Alignment

3.3 Face Recognition

After detection and alignment, the next step is to train the face recognition model. We chose to train the model with Caffe framework, a deep learning framework developed by Berkeley AI Research (https://caffe.berkeleyvision.org/). Following Wen et al. (2016), we use typical Convolutional Neural Networks with joint supervision of Softmax loss and Center loss. Face recognition use inter-class and intra-class variations to identify faces from pixels. External factors such as light, covers and expressions may cause the same person to appear differently. Inter-class variation is used to distinguish individuals. Intra-class variation is used to overlook these factors and to identify the same person. We can obtain inter-class dispension by Softmax and intral-class compactness by Center Loss. Figure 5 illustrates the network structure in Caffe. The convolutional layer is followed by the PReLU (Parametric rectified Linear Unit) (He et al., 2015) as it activation function (Cybenko & G. ,1989). PReLU is a type of ReLU (Jarrett et al., 2009) added parametric rectified. As illustrated in Figure 5, center loss is introduced in the feature layer output to reach the intra-class aggregation and the inter-class separation.

C: The convolution layer P: The max-pooling layer
LC: The local convolution layer FC: The fully connected layer

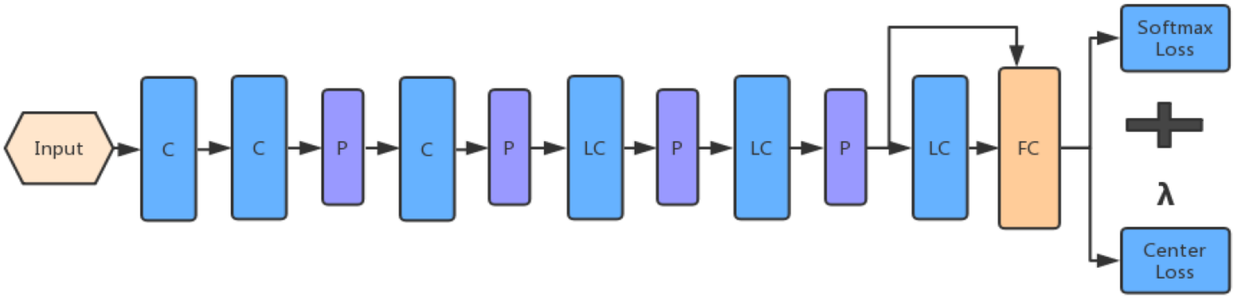


Figure 5. Face Recognition Network Structure Diagram

Softmax loss is calculated using formula (5)

$$L_s =-\sum_{i=1}^{m} log \frac{e^{w_{y_i}^T x_i+b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i+b_j}} \quad (5)$$

Xi - the ith deep feature in the d-dimensional space. Belongs to the yi class;
d - the dimension of the feature space;
*W - Fully connected layer parameter matrix. W={d*n}. d rows and n columns;*
Wj - the jth column of W;
M -batch size of mini- batch; N - Number of classes; b - bias
It can be seen that the denominator is all classes and the numerator is a single class.

Formula (6) represents Center loss calculation, where c_{yi} represents the central feature of the categories y_i, x_i represents the characteristics before fully connected layer. *m* represents the size of the mini-batch. We want to minimize the distance quadratic sum between sample's feature and the center of the feature.

$$L_c = \frac{1}{2}\sum_{i=1}^{m}\left|x_i - c_{y_i}\right|_2^2 \qquad (6)$$

In order to minimize the distance between the classes, we use the gradient descent method. Gradient formula as shown in formula (7).

$$\frac{\alpha L_c}{\partial x_i} = x_i - c_{y_i}$$

$$\Delta c_j = \frac{\sum_{i=1}^{m}\delta(y_i=j)\cdot(c_j-x_i)}{1+\sum_{i=1}^{m}\delta(y_i=j)} \qquad (7)$$

Finally, we combine softmax and center loss. λ is used to balance softmax loss with center loss. The larger the value, the greater the intra-class discrimination .

$$L = L_s + \lambda L_c \quad = -\sum_{i=1}^{m} log\frac{e^{w_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}} + \frac{\lambda}{2}\sum_{i=1}^{m}\left|x_i - c_{y_i}\right|_2^2 \qquad (8)$$

3.4 Similarity Calculation

The final step to find doppelgangers is image similarity calculation. Principal Component Analysis (PCA) is used to reduce feature dimensions (Sirovich et al., 1987). Cosine similarity is a popular similarity calculation that measures the cosine angle between two non-zero vectors of an inner product space. The smaller the angle, the more similar the two vectors are. Thus, a cosine similarity approaching 1 means the two vectors are extremely similar. Formula (9) is the cosine similarity calculation, where F1 and F2 represent extracted feature vectors from face recognition.

$$sim(F1, F2) = cos\theta = \frac{F1\cdot F2}{|F1|\cdot|F2|} \qquad (9)$$

4. Experiments

To evaluate the model, we conducted two experiments, the first experiment on a common dataset, the LFW Deep Funneled Images(http://vis-www.cs.umass.edu/lfw/), and the second experiment on a real-life a collection of over 3,000 students collected from ShanghaiTech.

4.1 Experiment 1: Validating Face Recognition using LFW Dataset

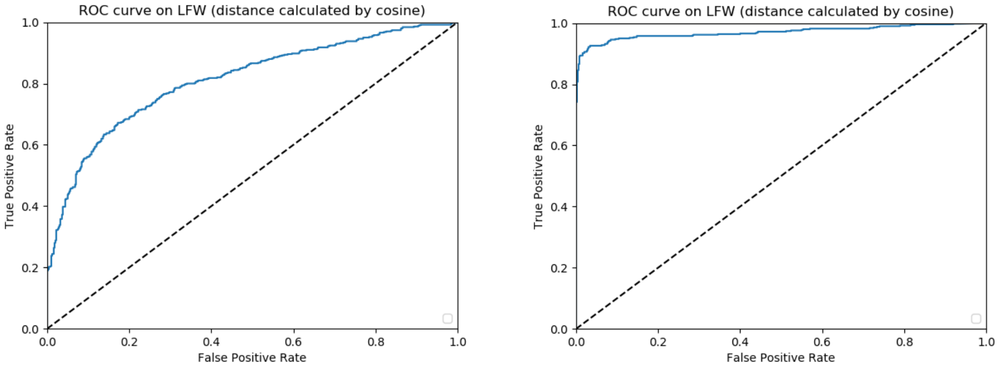
In the first experiment, we aim to validate our face recognition model, including the detection, alignment and recognition phases. We used a common dataset from LFW Deep Funneled Images (Huang et al., 2017). The dataset contains a total of 13,233 labeled pictures and is available at http://vis-www.cs.umass.edu/lfw/. The performance is reported using a classification algorithm which classifies multiple pictures of the same person. Notice that the data is only labeled with the same person for classification purpose. Although similarity algorithm cannot be directly validated here, the dataset is a good labeled source to validate face recognition algorithm. Furthermore, we examined the influence of face alignment in improving the recognition results. The alignment phase calibrates face position according to the five standard mean landmarks.

Table 1 illustrates the classification results with and without alignment adjustment. Figure 6 illustrates the ROC curve with alignment performed on only training dataset, and alignment performed on both training and texting dataset. It shows that alignment greatly improve the performance in terms of average precision, AP, EER and TPR (true positive rate). Average precision improved from 78.6% to 96.6% when alignment is properly performed.

Table 1. Validation Analysis of Alignment Performance

	AP (Average Prec.)	EER (equal error rate)	TP rate @0*	TP rate @001	TP rate @0001	TP rate @00001
Without Alignment	78.6	67.3	1.07	12.7	6.8	NA
With alignment for training data	82.5	78.8	19.9	31.3	26.5	NA
With alignment for training and testing data	96.8	97.2	71.7	92.3	89.4	NA

**TP Rate@0, 001, 0001, 00001 are true positive rate at false positive rate is equal or lesser than 0, 0.001, 0.0001 and 0.00001*



(a) (b)

Figure 6. (a) ROC curve with alignment performed for training only; (b) ROC curve with alignment performed for both training and testing

TO

4.2 Experiment 2: Finding Doppelgangers with Real-life Data

In experiment 2, we obtained a collection of over 3,611 student photos for our doppelganger mining. Following our proposed method, we performed MTCNN-based face detection, face alignment, and a CNN-based face recognition algorithm with softmax loss and center loss. Finally, cosine similarity was performed on all pair-wise images and similarity score above 0.8 were selected. This threshold gave us a total of 1,218 pairs of images, which accounts for 0.01% of total pair-wised sample as shown in Table 2. We found that five pairs with cosine score higher than 0.9 are actually from the same person, as illustrated in Figure 7. They submitted photos both during undergraduate and graduate studies. These five pairs were discarded. The fact that the algorithm can successfully identify the same person shows that our face recognition and similarity measure algorithms can be used in other applications with a higher threshold.

We then manually checked the 1,218 pairs of images and group them into two categories: pairs that are highly alike, or doppelgangers; and pairs that are somewhat alike, not real doppelgangers.

The results are concluded to Figure 8. However, we did not perform a formal tagging to generate a gold standard from the dataset. This manual check was explorative.

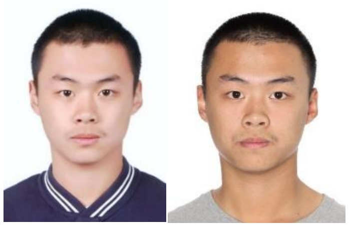


Figure 7. Pictures from the same student with cosine similarity score of 0.905

Table 2. Similarity Score Distribution and Doppelganger Selection

Total number of images	3,611	
Total number of image pairs	6,517,855	
# of images with cosine score > 0.8	**1,218**	**0.01% of total sample**
# of images with cosine score > 0.9	5	discarded

Cosine Similarity Score

Similarity > 0.9 → **same person**

0.8 < Similarity < 0.9 → highly alike person → **True Doppelgangers**

0.7 < Similarity < 0.8 → somewhat alike person → Alike but **not Doppelgangers**

Figure 8. Doppelganger Identification

Figure 9 shows 6 pairs of top-ranked face photos with cosine similarity greater than 0.8 and are judged by human as highly alike pairs. These student pairs are considered true doppelgangers in our study. We plan to organize a photo shooting session with these students to mimic Brunel photography's exhibition.

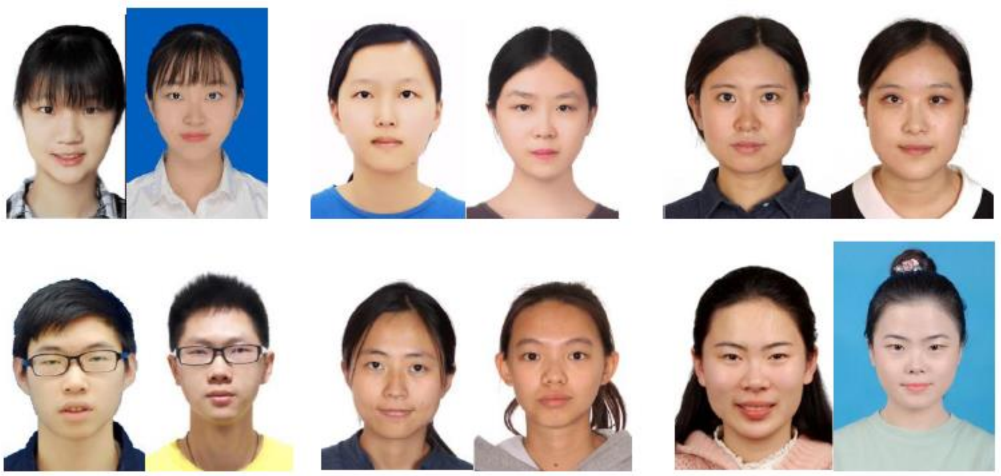


Figure 9. Real-life Doppelgangers on campus: top-ranked face pairs (0.7< cosine similarity <0.85 combination)

For the somewhat alike but not real doppelganger group, we further investigated the possible reasons of their high cosine similarity score. We identified the following areas that could cause high cosine similarity while the two faces are not highly alike.

(1) The model is sensitive to facial expression. As shown in Figure 10, the two people smile the same way have a higher chance of being matched with cosine similarity > 0.8.

(2) The model is sensitive to brightened skin, especially with pictures that are post-processed. For example, in Figure 11, both skin colors were brightened, and facial flaws were blurred. Although they are shown to have high cosine similarity, the two people are less alike. We believe during PCA phase, such skin processing could largely affect the performance of how dimensions are mapped and reduced.

(3) The model is sensitive to facial lines (wrinkles). As shown in Figure 12, both faces have deep smile lines and slightly raised mouths. These two features become dominant features. This also means, when people age, the recognition algorithm may fail to recognize them with wrinkles. The sensitivity level can be adjusted during the learning process.

(4) The model is sensitive to accessories on faces. As shown in Figure 13, two people wearing similar eyeglasses are selected as top-ranked results. The shape and color of glasses are taken into the model as important features in similarity measure.

(5) There is also a dilemma with alignment. We found one face image that has very close landmarks with the mean standard landmarks. As a result, the model identified 12 other face images that are lookalikes. We consider this to be a false positive example.

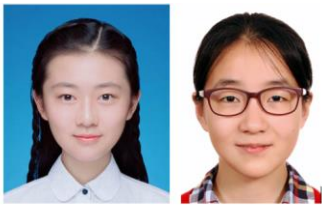


Figure 10. Matching results affected by facial expressions such as smile styles

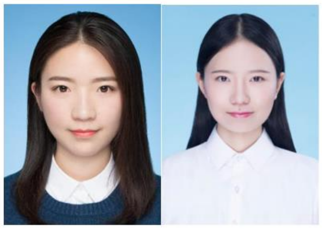


Figure 11. Matching results affected by post-processing such as skin brightening

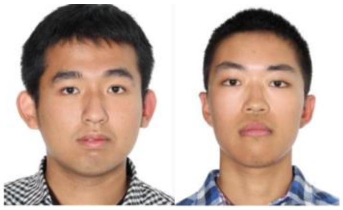


Figure 12. Matching results affected by facial lines

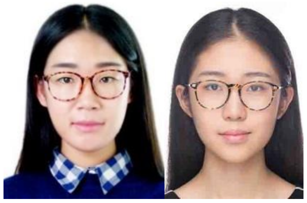


Figure 13. Matching results affected by wearing eyeglasses

5. Discussions

Although the focus of this study is on doppelganger mining, our algorithm can be extended to other applications in e-business. However, the threshold for similarity measure needs to be tested. Figure 14 shows two images of the same person Gigi Hadid. Picture on the left is a picture used in a store in Taobao platform. Picture on the right is a runway picture of Hadid. The similarity score here is 0.715. Our testing results on LFW image database also shows that on average, pictures of the same persons' similarity score are 0.5-0.8. Our study is based on students' ID photos and the similarity threshold should be set higher.

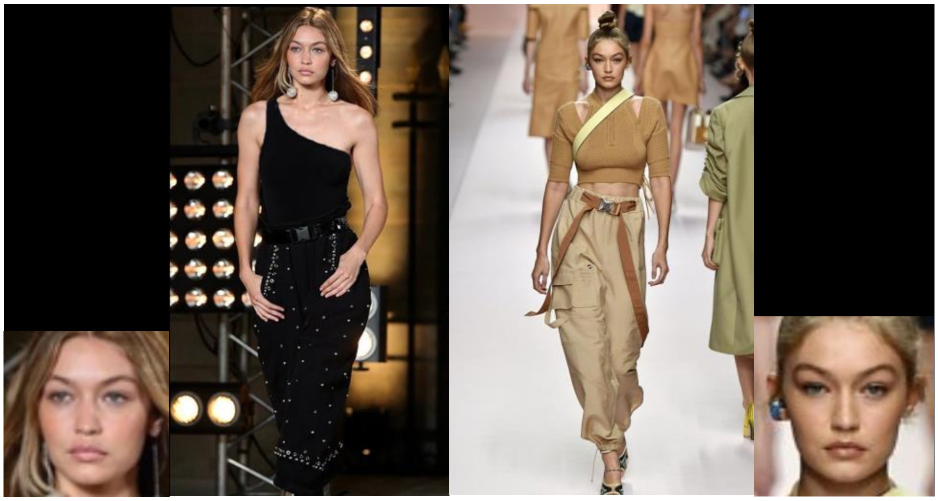


Figure 14. Catwalk Picture Similarity Calculation (with cosine similarity = 0.715)

4 Conclusions

This study aims to perform doppelganger mining by using MTCNN-based face detection and CNN-based face recognition algorithm. MTCNN is an effective method for face detection. Both softmax loss and center loss functions are used in face recognition to identify the best face features. The model is trained on Caffe framework. Using a dataset of over 3000 face images, we are able to identify pairs of doppelgangers that do look alike each other. We also found that the model is sensitive to facial expression, skin brightness, skin lines and accessories on face.

In the future, we would like to exhibit the results of this study in the form of pairs of pictures, similar to Brunel Photography's Exhibition. This will inspire students of non-CS major to explore deep learning and image recognition technology. Furthermore, it will encourage collaboration across-disciplines on campus. In terms of e-business application, the similarity comparison algorithm shows great potential in detecting *violations of portrait right.*

In the future, we plan to further improve our model to address the current four problems we observe, and also apply the model to more application domains.

References

Chase Jarvis Photography (2018). Available at https://www.chasejarvis.com/photos/

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems, 2(4), 303-314.

Gary B. Huang, Manu Ramesh, Tamara Berg, & Erik Learned-Miller. (2007, October). Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49.

Georghiades, A. S., Belhumeur, P. N., & Kriegman, D. J. (2001). From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis & Machine Intelligence, (6), 643-660.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (pp. 1026-1034).

Jarrett, K., Kavukcuoglu, K., & LeCun, Y. (2009, September). What is the best multi-stage architecture for object recognition?. In 2009 IEEE 12th international conference on computer vision (pp. 2146-2153). IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

Logie, R. H., Baddeley, A. D., & Woodhead, M. M. (1987). Face recognition, pose and ecological validity. Applied Cognitive Psychology, 1(1), 53-69.

Nguyen, H. V., & Bai, L. (2010, November). Cosine similarity metric learning for face verification. In Asian conference on computer vision (pp. 709-720). Springer, Berlin, Heidelberg.

Shen, L., & Bai, L. (2006). A review on Gabor wavelets for face recognition. Pattern analysis and applications, 9(2-3), 273-292.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. Josa a, 4(3), 519-524.

Smirnov, E., Melnikov, A., Novoselov, S., Luckyanets, E., & Lavrentyeva, G. (2017). Doppelganger mining for face representation learning. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1916-1923).

Sun, Y., Wang, X., & Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1891-1898).

Suri, P. K., Walia, E., & Verma, E. A. (2011, May). Novel face detection using Gabor filter bank with variable threshold. In 2011 IEEE 3rd International Conference on Communication Software and Networks (pp. 715-720). IEEE.

Sim, T., Baker, S., & Bsat, M. (2002, May). The CMU pose, illumination, and expression (PIE) database. In Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition (pp. 53-58). IEEE.

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708).

Time (2018). Available at https://time.com/5435683/artificial-intelligence-painting-christies/

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. CVPR (1), 1(511-518), 3.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. International journal of computer vision, 57(2), 137-154.

Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016, October). A discriminative feature learning approach for deep face recognition. In European conference on computer vision (pp. 499-515). Springer, Cham.

Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2014, September). Facial landmark detection by deep multi-task learning. In European conference on computer vision (pp. 94-108). Springer, Cham.

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10), 1499-1503.